Generalized Orlicz Spaces and Wasserstein Distances for Convex-Concave Scale Functions

Karl-Theodor Sturm

Abstract

Given a strictly increasing, continuous function $\vartheta : \mathbb{R}_+ \to \mathbb{R}_+$, based on the cost functional $\int_{X\times X} \vartheta\left(d(x,y)\right) dq(x,y)$, we define the L^ϑ-Wasserstein distance $W_\vartheta(\mu,\nu)$ between probability measures μ,ν on some metric space (X,d). The function ϑ will be assumed to admit a representation $\vartheta=\varphi\circ\psi$ as a composition of a convex and a concave function φ and ψ, resp. Besides convex functions and concave functions this includes all $\mathcal{C}^2$ functions.

For such functions ϑ we extend the concept of Orlicz spaces, defining the metric space $L^\vartheta(X,m)$ of measurable functions $f : X \to \mathbb{R}$ such that, for instance,

$$d_\vartheta(f,g) \le 1 \quad \Longleftrightarrow \quad \int_X \vartheta(|f(x)-g(x)|)\, d\mu(x) \le 1.$$

1 Convex-Concave Compositions

Throughout this paper, ϑ will be a strictly increasing, continuous function from $\mathbb{R}_+$ to $\mathbb{R}_+$ with $\vartheta(0)=0$.

Definition 1.1. *ϑ will be called ccc function ("convex-concave composition") iff there exist two strictly increasing continuous functions $\varphi,\psi : \mathbb{R}_+ \to \mathbb{R}_+$ with $\varphi(0)=\psi(0)=0$ s.t. φ is convex, ψ is concave and*

$$\vartheta = \varphi\circ\psi.$$

The pair (φ,ψ) will be called convex-concave factorization of ϑ.

The factorization is called minimal (or non-redundant) if for any other factorization $(\tilde\varphi,\tilde\psi)$ the function $\varphi^{-1}\circ\tilde\varphi$ is convex.

Two minimal factorizations of a given function ϑ differ only by a linear change of variables. Indeed, if $\varphi^{-1}\circ\tilde\varphi$ is convex and also $\tilde\varphi^{-1}\circ\varphi$ is convex then there exists a $\lambda\in(0,\infty)$ s.t. $\tilde\varphi(t)=\varphi(\lambda t)$ and $\tilde\psi(t)=\frac{1}{\lambda}\psi(t)$.

For each convex, concave or ccc function $f:\mathbb{R}_+\to\mathbb{R}_+$ put $f'(t) := f'(t+) := \lim_{h\searrow 0}\frac{1}{h}\left[f(t+h)-f(t)\right]$.

Lemma 1.2. *(i) For any ccc function ϑ, the function $\log\vartheta'$ is locally of bounded variation and the distribution $(\log\vartheta')'$ defines a signed Radon measure on $(0,\infty)$, henceforth denoted by $d(\log\vartheta')$.*

(ii) A pair (φ,ψ) of strictly increasing convex or concave, resp., continuous functions with $\varphi(0)=\psi(0)=0$ is a factorization of ϑ iff

$$d(\log\vartheta') = \psi_*^{-1} d(\log\varphi') + d(\log\psi') \tag{1}$$

in the sense of signed Radon measures.

(iii) The factorization (φ,ψ) is minimal iff for any other factorization $(\tilde\varphi,\tilde\psi)$

$$-d(\log\psi') \le -d(\log\tilde\psi')$$

in the sense of nonnegative Radon measures on $(0,\infty)$.

(iv) Every ccc function ϑ admits a minimal factorization $(\check\vartheta,\hat\vartheta)$ given by $\check\vartheta := \vartheta\circ\hat\vartheta^{-1}$ and

$$\hat\vartheta(x) := \int_0^x \exp\left(-\int_1^y d\nu_-(z)\right) dy$$

where $d\nu_-(z)$ denotes the negative part of the Radon measure $d\nu(z)=d(\log\vartheta')(z)$.

Proof. (i), (ii): The chain rule for convex/concave functions yields

$$\vartheta'(t) = \varphi'(\psi(t)) \cdot \psi'(t)$$

for each factorization (φ, ψ) of a ccc function ϑ. Taking logarithms it implies that $\log \vartheta'$ locally is a BV function (as a difference of two increasing functions) and, hence, that the associated Radon measures satisfy

$$\begin{aligned} d(\log \vartheta') &= d(\log \varphi' \circ \psi) + d(\log \psi') \\ &= \psi_*^{-1} d(\log \varphi') + d(\log \psi'). \end{aligned}$$

(iii): The factorization (φ, ψ) is minimal if and only if for any other factorization $(\tilde{\varphi}, \tilde{\psi})$ the function $u = \varphi^{-1} \circ \varphi = \psi \circ \tilde{\psi}^{-1}$ is convex. Since $\log \psi' = \log u'(\tilde{\psi}) + \log \tilde{\psi}'$, the latter is equivalent to

$$d(\log \psi') \geq d(\log \tilde{\psi}')$$

which is the claim.

(iv): Define $\hat{\vartheta}$ as above. It remains to verify that $\hat{\vartheta} < \infty$. Let (φ, ψ) be any convex-concave factorization of ϑ. Without restriction assume $\psi'(1) = 1$. Then the Hahn decomposition of (1) yields

$$d\nu_- \leq -d(\log \psi'). \tag{2}$$

Hence, for all $0 \leq x \leq 1$

$$\begin{aligned} 0 \leq \hat{\vartheta}(x) &= \int_0^x \exp\left(\int_y^1 d\nu_-(z)\right) dy \\ &\leq \int_0^x \exp\left(-\int_y^1 d(\log \psi')(z)\right) dy = \psi(x) < \infty. \end{aligned}$$

This already implies that $\hat{\vartheta}$ is finite, strictly increasing and continuous on $[0, \infty)$. (For instance, for $x > 1$ it follows $\hat{\vartheta}(x) \leq \hat{\vartheta}(1) + x - 1$.) Moreover, one easily verifies that $\hat{\vartheta}$ is concave.

Since ν_+, ν_- are the minimal nonnegative measures in the ('Hahn' or 'Jordan') decomposition of $\nu = \nu_+ - \nu_-$, it follows that $(\check{\vartheta}, \hat{\vartheta})$ is a minimal cc decomposition of ϑ. □

Examples 1.3.

- *Each convex function ϑ is a ccc function. A minimal factorization is given by (ϑ, Id).*
- *Each concave function ϑ is a ccc function. A minimal factorization is given by (Id, ϑ).*
- *Each $\mathcal{C}^2$ function ϑ with $\vartheta'(0+) > 0$ is a ccc function. The minimal factorization is given by*

  $$\hat{\vartheta}(x) := \int_0^x \exp\left(\int_1^y \frac{\vartheta''(z) \wedge 0}{\vartheta'(z)} dz\right) dy$$

  and $\check{\vartheta} := \vartheta \circ \hat{\vartheta}^{-1}$. (The condition $\vartheta'(0+) > 0$ can be replaced by the strictly weaker requirement that the previous integral defining $\hat{\vartheta}$ is finite.)

2 The Metric Space $L^\vartheta(X, \mu)$

Let (X, Ξ, μ) be a σ-finite measure space and (φ, ψ) a minimal ccc factorization of a given function ϑ. Then $L^\vartheta(X, \mu)$ will denote the space of all measurable functions $f : X \to \mathbb{R}$ such that

$$\int_X \varphi\left(\frac{1}{t}\psi(|f|)\right) d\mu < \infty$$

for some $t \in (0, \infty)$ where as usual functions which agree almost everywhere are identified. Note that – due to the fact that $r \mapsto \varphi(r)$ for large r grows at least linearly – the previous condition is equivalent to the condition $\int_X \varphi\left(\frac{1}{t}\psi(|f|)\right) d\mu \leq 1$ for some $t \in (0, \infty)$.

Theorem 2.1. *$L^\vartheta(X, \mu)$ is a complete metric space with the metric*

$$d_\vartheta(f, g) = \inf\left\{t \in (0, \infty) : \int_X \varphi\left(\frac{1}{t}\psi(|f - g|)\right) d\mu \leq 1\right\}.$$

The definition of this metric does not depend on the choice of the minimal ccc factorization of the function ϑ. However, choosing an arbitrary convex-concave factorization of ϑ might change the value of d_ϑ.

Note that always $d_\vartheta(f,g) = d_\vartheta(f-g,0)$.

Proof. Let $f,g,h \in L^\vartheta(X,\mu)$ be given and choose $r,s>0$ with $d_\vartheta(f,g)<r$ and $d_\vartheta(g,h)<s$. The latter implies

$$\int_X \varphi\left(\frac{1}{r}\psi(|f-g|)\right)\,d\mu \le 1, \qquad \int_X \varphi\left(\frac{1}{s}\psi(|g-h|)\right)\,d\mu \le 1.$$

Concavity of ψ yields $\psi(|f-h|) \le \psi(|f-g|) + \psi(|g-h|)$. Put $t = r+s$. Then convexity of φ implies

$$\varphi\left(\frac{1}{t}\psi(|f-h|)\right) \le \varphi\left(\frac{r}{t}\cdot\frac{\psi(|f-g|)}{r} + \frac{s}{t}\cdot\frac{\psi(|g-h|)}{s}\right) \le \frac{r}{t}\cdot\varphi\left(\frac{\psi(|f-g|)}{r}\right) + \frac{s}{t}\cdot\varphi\left(\frac{\psi(|g-h|)}{s}\right).$$

Hence,

$$\int_X \varphi\left(\frac{1}{t}\psi(|f-h|)\right)d\mu \le \frac{r}{t}\cdot\int_X \varphi\left(\frac{\psi(|f-g|)}{r}\right)d\mu + \frac{s}{t}\cdot\int_X \varphi\left(\frac{\psi(|g-h|)}{s}\right)d\mu \le \frac{r}{t}\cdot 1 + \frac{s}{t}\cdot 1 = 1$$

and thus $d_\vartheta(f,h) \le t$. This proves that $d_\vartheta(f,h) \le d_\vartheta(f,g) + d_\vartheta(g,h)$.

In order to prove the completeness of the metric, let $(f_n)_n$ be a Cauchy sequence in L^ϑ. Then $d_\vartheta(f_n,f_m) < \epsilon_n$ for all n,m with $m \ge n$ and suitable $\epsilon_n \searrow 0$. Choose an increasing sequence of measurable sets X_k, $k\in\mathbb{N}$, with $\mu(X_k)<\infty$ and $\cup_k X_k = X$. Then

$$\int_{X_k} \varphi\left(\frac{1}{\epsilon_n}\psi(|f_n - f_m|)\right)d\mu \le 1$$

for all k,m,n with $m\ge n$. Jensen's inequality implies

$$\varphi\left(\frac{1}{\mu(X_k)}\int_{X_k}\frac{1}{\epsilon_n}\psi(|f_n-f_m|)\,d\mu\right) \le \frac{1}{\mu(X_k)}$$

and thus

$$\int_{X_k} |\psi(f_n) - \psi(f_m)|\,d\mu \le \epsilon_n \cdot \mu(X_k)\cdot\varphi^{-1}\left(\frac{1}{\mu(X_k)}\right).$$

In other words, $(\psi(f_n))_n$ is a Cauchy sequence in $L^1(X_k,\mu)$. It follows that it has a subsequence $(\psi(f_{ni}))_i$ which converges μ-almost everywhere on X_k. In particular, $(f_{ni})_i$ converges almost everywhere on X_k towards some limiting function f (which easily is shown to be independent of k).

Finally, Fatou's lemma now implies

$$\int_{X_k}\varphi\left(\frac{1}{\epsilon_n}\psi(|f_n-f|)\right)d\mu \le \liminf_{m\to\infty}\int_{X_k}\varphi\left(\frac{1}{\epsilon_n}\psi(|f_n-f_m|)\right)d\mu \le 1$$

for each k and $n\in\mathbb{N}$. Hence,

$$\int_X \varphi\left(\frac{1}{\epsilon_n}\psi(|f_n-f|)\right)d\mu \le 1,$$

that is,

$$d_\vartheta(f_n,f) \le \epsilon_n$$

which proves the claim.

Finally, it remains to verify that

$$d_\vartheta(f,g) = 0 \quad\Longleftrightarrow\quad f = g\ \mu\text{-a.e. on } X.$$

The implication $\Leftarrow$ is trivial. For the reverse implication, we may argue as in the previous completeness proof: $d_\vartheta(f,g)=0$ will yield $\int_{X_k}\varphi\left(\frac{1}{t}\psi(|f-g|)\right)\,d\mu \le 1$ for all $k\in\mathbb{N}$ and all $t>0$ which in turn implies $\int_{X_k}|\psi(f)-\psi(g)|\,d\mu = 0$. The latter proves $f=g$ μ-a.e. on X which is the claim. ∎

Examples 2.2. *If $\vartheta(r) = r^p$ for some $p\in(0,\infty)$ then*

$$d_\vartheta(f,g) = \left(\int_X |f-g|^p\,d\mu\right)^{1/p^*}$$

with $p^* := p$ if $p\ge 1$ and $p^* := 1$ if $p \le 1$.

Proposition 2.3. *(i) If ϑ is convex then $\|f\|_{L^\vartheta(X,\mu)} := d_\vartheta(f,0)$ is indeed a norm and $L^\vartheta(X,\mu)$ is a Banach space, called Orlicz space. The norm is called Luxemburg norm.*

(ii) If ϑ is concave then

$$d_\vartheta(f,g) = \int_X \vartheta(|f-g|)\,d\mu \geq \|\vartheta(f)-\vartheta(g)\|_{L^1(X,\mu)}.$$

(iii) For general ccc function $\vartheta = \varphi \circ \psi$

$$d_\vartheta(f,g) = \|\psi(|f-g|)\|_{L^\varphi(X,\mu)}.$$

(iv) If $\mu(M)=1$ then for each strictly increasing, convex function $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ with $\Phi^{-1}(1)=1$

$$d_{\Phi\circ\vartheta}(f,g) \geq d_\vartheta(f,g)$$

("Jensen's inequality").

Proof. (i) If $\psi(r) = cr$ then obviously $d_\vartheta(tf,0) = t \cdot d_\vartheta(f,0)$. See also standard literature [2].

(ii) Concavity of ϑ implies $\vartheta(|f-g|) \geq |\vartheta(f)-\vartheta(g)|$.

(iv) Assume that $d_{\Phi\circ\vartheta}(f,g) < t$ for some $t \in (0,\infty)$. It implies

$$\int_X \Phi\left(\varphi\left(\frac{1}{t}\psi(|f-g|)\right)\right)\,d\mu \leq 1.$$

Classical Jensen inequality for integrals yields

$$\Phi\left(\int_X \varphi\left(\frac{1}{t}\psi(|f-g|)\right)\,d\mu\right) \leq 1$$

which – due to the fact that $\Phi^{-1}(1)=1$ – in turn implies $d_\vartheta(f,g) \leq t$. ☐

3 The L^ϑ-Wasserstein Space

Let (X,d) be a complete separable metric space and ϑ a ccc function with minimal factorization (φ,ψ). The *L^ϑ-Wasserstein space* $\mathcal{P}_\vartheta(X)$ is defined as the space of all probability measures μ on X – equipped with its Borel σ-field – s.t.

$$\int_X \varphi\left(\frac{1}{t}\psi(d(x,y))\right)\,d\mu(x) < \infty$$

for some $y \in X$ and some $t \in (0,\infty)$. The L^ϑ-Wasserstein distance of two probability measures $\mu,\nu \in \mathcal{P}_\vartheta(X)$ is defined as

$$W_\vartheta(\mu,\nu) \quad = \quad \inf\left\{t>0 : \inf_{q\in\Pi(\mu,\nu)} \int_{X\times X} \varphi\left(\frac{1}{t}\psi(d(x,y))\right)\,dq(x,y) \leq 1\right\}$$

where $\Pi(\mu,\nu)$ denotes the set of all couplings of μ and ν, i.e. the set of all probability measures q on $X\times X$ s.t. $q(A\times X) = \mu(A), q(X\times A) = \nu(A)$ for all Borel sets $A \subset X$.

Given two probability measures $\mu,\nu \in \mathcal{P}_\vartheta(X)$, a coupling q of them is called *optimal* iff

$$\int_{X\times X} \varphi\left(\frac{1}{w}\psi(d(x,y))\right)\,dq(x,y) \leq 1$$

for $w := W_\vartheta(\mu,\nu)$.

Proposition 3.1. *For each pair of probability measures $\mu,\nu \in \mathcal{P}_\vartheta(X)$ there exists an optimal coupling q.*

Proof. For $t \in (0,\infty)$ define the cost function $c_t(x,y) = \varphi(\frac{1}{t}\psi(d(x,y)))$. Note that $t \mapsto c_t(x,y)$ is continuous and decreasing.

Given μ,ν s.t. $w := W_\vartheta(\mu,\nu) < \infty$. Then for all $t > w$ the measures μ and ν have finite c_t-transportation costs. More precisely,

$$\inf_{q\in\Pi(\mu,\nu)} \int_{X\times X} c_t(x,y)\,dq(x,y) \leq 1.$$

Hence, there exists $q_n \in \Pi(\mu, \nu)$ s.t.

$$\int_{X \times X} c_{w+\frac{1}{n}}(x, y)\, dq_n(x, y) \leq 1 + \frac{1}{n}.$$

In particular, $\int_{X\times X} c_{w+1}(x, y))\, dq_n(x, y) \leq 2$ for all $n \in \mathbb{N}$. Hence, the family $(q_n)_n$ is tight ([3], Lemma 4.4). Therefore, there exists a converging subsequence $(q_{n_k})_k$ with limit $q \in \Pi(\mu, \nu)$ satisfying

$$\int_{X \times X} c_{w+\frac{1}{n}}(x, y)\, dq(x, y) \leq 1 + \frac{1}{n}$$

for all n ([3], Lemma 4.3) and thus

$$\int_{X \times X} c_w(x, y)\, dq(x, y) \leq 1.$$

□

Proposition 3.2. *W_ϑ is a complete metric on $\mathcal{P}_\vartheta(X)$.*

The triangle inequality for W_ϑ is valid not only on $\mathcal{P}_\vartheta(X)$ but on the whole space $\mathcal{P}(X)$ of probability measures on X. The triangle inequality implies that $W_\vartheta(\mu, \nu) < \infty$ for all $\mu, \nu \in \mathcal{P}_\vartheta(X)$.

Proof. Given three probability measures μ_1, μ_2, μ_3 on X and numbers r, s with $W_\vartheta(\mu_1, \mu_2) < r$ and $W_\vartheta(\mu_2, \mu_3) < s$. Then there exist a coupling q_{12} of μ_1 and μ_2 and a coupling q_{23} of μ_2 and μ_3 s.t.

$$\int \varphi\left(\frac{1}{r}\psi \circ d\right) dq_{12} \leq 1, \qquad \int \varphi\left(\frac{1}{s}\psi \circ d\right) dq_{23} \leq 1.$$

Let q_{123} be the gluing of the two couplings q_{12} and q_{23}, see e.g. [1], Lemma 11.8.3. That is, q_{123} is a probability measure on $X \times X \times X$ s.t. the projection onto the first two factors coincides with q_{12} and the projection onto the last two factors coincides with q_{23}. Let q_{13} denote the projection of q_{123} onto the first and third factor. In particular, this will be a coupling of μ_1 and μ_3. Then for $t := r + s$

$$\begin{aligned}
&\int_{X \times X} \varphi\left(\frac{1}{t}\psi(d(x, z))\right) dq_{13}(x, z) \\
\leq\; & \int_{X\times X\times X} \varphi\left(\frac{1}{t}\psi(d(x, y) + d(y, z))\right) dq_{123}(x, y, z) \\
\leq\; & \int_{X\times X\times X} \varphi\left(\frac{r}{t}\frac{\psi(d(x, y))}{r} + \frac{s}{t}\frac{\psi(d(y, z))}{s}\right) dq_{123}(x, y, z) \\
\leq\; & \frac{r}{t}\int_{X\times X\times X} \varphi\left(\frac{\psi(d(x, y))}{r}\right) dq_{123}(x, y, z) + \frac{s}{t}\int_{X\times X\times X} \varphi\left(\frac{\psi(d(y, z))}{s}\right) dq_{123}(x, y, z) \\
\leq\; & \frac{r}{t} \cdot 1 + \frac{s}{t} \cdot 1 = 1.
\end{aligned}$$

Hence, $W_\vartheta(\mu_1, \mu_3) \leq t$. This proves the triangle inequality.

To prove completeness, assume that $(\mu_k)_k$ is a W_ϑ-Cauchy sequence, say $W_\vartheta(\mu_n, \mu_k) \leq t_n$ for all $k \geq n$ with $t_n \to 0$ as $n \to \infty$. Then there exist couplings $q_{n,k}$ of μ_n and μ_k s.t.

$$\int \varphi\left(\frac{1}{t_n}\psi(d(x, y))\right) dq_{n,k}(x, y) \leq 1. \tag{3}$$

Jensen's inequality implies

$$\int \tilde{d}(x, y)\, dq_{n,k}(x, y) \leq t_n \cdot \varphi^{-1}(1)$$

with $\tilde{d}(x, y) := \psi(d(x, y))$. The latter is a complete metric on X with the same topology as d. That is, $(\mu_k)_k$ is a Cauchy sequence w.r.t. the L^1-Wasserstein distance on $\mathcal{P}(X, \tilde{d})$. Because of completeness of $\mathcal{P}_1(X, \tilde{d})$, we thus obtian an accumulation point μ and a converging subsequence $(\mu_{k_i})_i$. According to [3], Lemma 4.4, this also yields an accumulation point q_n of the sequence $(q_{n,k_i})_i$. Continuity of the involved cost functions – together with Fatou's lemma – allows to pass to the limit in (3) to derive

$$\int \varphi\left(\frac{1}{t_n}\psi(d(x, y))\right) dq_n(x, y) \leq 1$$

which proves that $W_\vartheta(\mu,\mu_n) \le t_n \to 0$ as $n \to \infty$.

With a similar argument, one verifies that $W_\vartheta(\mu,\nu) = 0$ if and only if $\mu = \nu$. $\square$

Remark 3.3. *For each pair of probability measures μ, ν on X*

$$W_\vartheta(\mu,\nu) \le 1 \quad \Longleftrightarrow \quad \inf_{q\in\Pi(\mu,\nu)} \int_{X\times X} \vartheta(d(x,y))\, dq(x,y) \le 1.$$

References

[1] R.M. Dudley: *Real analysis and probability.* Cambridge Univ Pr, 2002.

[2] M.M. Rao, Z.D. Ren: *Theory of Orlicz Spaces.* Pure and Applied Mathematics, Marcel Dekker (1991).

[3] C. Villani: *Optimal Transport, old and new.* Grundlehren der mathematischen Wissenschaften 338 (2009), Springer Berlin · Heidelberg.